

# Semantic Web and Information Systems: An Agenda Based on Discourse with Community Leaders

Miltiadis D. Lytras, Athens University of Economics and Business, Greece

## ABSTRACT

When last January, AIS approved our proposal for a new SIG on Semantic Web and Information Systems ([www.sigs.emis.org](http://www.sigs.emis.org)), I never thought that we could gain the support of so many renowned academics and practitioners. Moreover, I couldn't imagine that all these people would be so excited concerning knowledge sharing and community building around the Semantic Web and its catalytic influence on our traditional perceptions of expressing and exploiting meaning through tools, services, and applications. Having already interviewed four key people for the evolution of the Semantic Web, we decided instead to provide a simple editorial to sketch the open Semantic Web and information systems research agenda.

## INTRODUCTION

*Semantic Web = Semantics + Web* — a simple equation, yet so many ways to interpret it. I feel blessed that I had the opportunity and the honor to find guidance, vision, and different views from all the leaders of the Semantic Web — namely (and in order of time) Amit Sheth, James Hendler, Chris Bussler, and Eric Miller. Every time I finished an interview, I felt that these people contributed to all of us by stipulating our energy to

be part of a great and silent revolution, the one of the Semantic Web (SW). I will try in the next few pages to put together their thoughts and ideas with my questioning. My ultimate objective in the end is to interpret the initial equation. Semantics is the first, the most important part of the Semantic Web.

According to Sheth:

*“Semantics has long been recognized to be very important in IS, Databases, AI, Linguistics and many other fields. From the IS/DB perspective, I remember talking about ‘So Far (Schematically) yet So Near (Semantically)’ in 1992, but lots of smarter people have talked about semantics for some time. More recently however, two things have happened—one positive, one potentially not so positive. The positive thing is that we have now been able to engineer semantic technology that supports large-scale semantic applications, and use large populated ontologies to provide semantic underpinning. At the same time a questionable development is a rather overwhelming importance attached to ‘formal semantics.’” (Note: Sheth does not argue against the importance on formal semantics, he merely questions sole or overwhelming reliance on it since, for semantics, one needs to bridge a gap between humans, real world domain knowledge and the machines, and the formal representation works adequately well only for the machines.)*

In this inaugural issue, Amit Sheth, Cartic Ramakrishnan, and Christopher Thomas, all from the University of Georgia, pro-

vide an excellent discussion on “Semantics for the Semantic Web: The Implicit, the Formal, and the Powerful.” Considering the role of semantics in a number of research areas in computer science, they organize semantics in three forms — implicit, formal, and powerful — and explore their roles in enabling some of the key capabilities related to the Semantic Web. The central message of this article is that building the Semantic Web purely on description logics will artificially limit its potential, and that we will need to both exploit well-known techniques that support implicit semantics, and develop more powerful semantic techniques. This article is surely an excellent starting point for everyone interested in SW. Please note that since this first issue consists of invited papers rather than refereed, I have taken the liberty of asking the EIC to contribute.

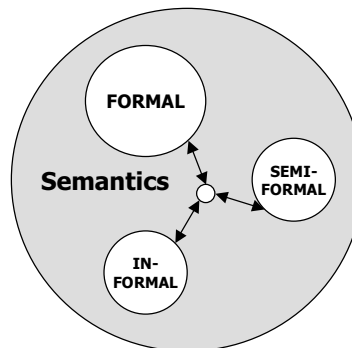
If you ask a newcomer in SW the first thing he would like to know, likely his response will address the impact of a Semantic Web on services and applications. It is the key issue for promoting the visibility and the worthiness of SW. The starting point of such questioning is always the same: “What is the difference in comparison to the WWW that we all know?”

Eric Miller set an interesting “framework” for understanding the evolution of the Semantic Web in relation to the WWW evolution:

*“If we think back to the phases associated with Web deployment: 1. The Web was born at CERN, 2. Was first picked up by high-energy physicists, 3. Then by academia at large, 4. Then by small businesses and start-ups, 5. Big business came only later! I’d suggest the Semantic Web is now at #4, and very quickly moving to #5.”*

I believe this is only the half story. And of course Miller just pointed out the critical aspect: the Semantic Web is not an initiative hermetically tight to close academic clubs.

Figure 1. Semantics types



“Semantic Web is here to stay,” Sheth says with emphasis.

So from this perspective I totally agree with Sheth:

*“More exiting and important goals of the next generation Web research are improving the human experience and enriching the living, and I can now see a possibility of a major shift from focus on computing to improving human experience — not only with better ability to use heterogeneous content and apply knowledge, but also to incorporate perception and pervasive computing.”*

This ultimate objective reveals the key argumentation for technology adoption. We need technologies as means for improving our lives and expanding our frontiers towards the common wealth. But his statement hides a lot of engineering.

Chris Bussler provided a very interesting insight to this issue:

*“When you look behind the scenes, and study what information systems infrastructure has to be put in place and maintained in order to provide that level of services, be it for customers or businesses, the state of affairs can be improved quite a bit from a technological side, especially the semantics side of it. Too many glitches*

*happen due to missing semantic underpinning. Once semantics-based technologies are available, the situation for customers and businesses can be advanced a lot beyond the current state, too, in addition to overcoming today's problems."*

The emphasis on today's problems and on the vision for tomorrow incorporates the two major pillars of the so-called knowledge society. Sheth provided one more excellent insight with a philosophic flavor:

*"...It is fairly certain that nothing we are seeing is a utopia. We all have [the] tendency to get unduly excited with every new trend and fad, and after a field matures, we find out that instead of them being [a] major life-changing technology or science, they are a step towards a continuing evolution."*

This is the Semantic Web for me also. We do not have to underestimate the social context, not even the societal inquiries for new services.

Danny Ayer contributed in *AIS SIGSEMIS Bulletin* a very interesting article. I found his positions quite informing. His main point is that the current Web has many inefficiencies and characteristics that limit its value. Ayers' excellent description in Figure 2 sets an interesting context for revealing the required elements of the Semantic Web. In the next section we try to outline the key re-

search issues of the Semantic Web research agenda.

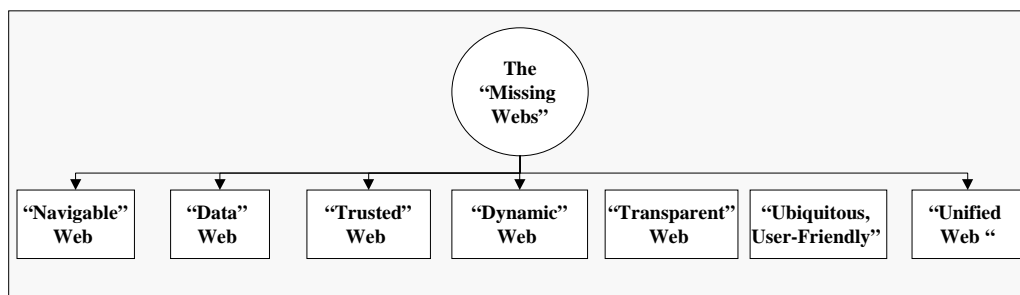
## Navigable Web

There is a huge amount of information on the Web, but it is of limited use without it being possible to access that information with ease. Compared to traditional systems the current Web is closer to a file system than a relational database. We have a means of storing and labeling the documents, what we don't have is any built-in technique for indexing and searching them. Catalogue-styled portals do help, and search engines like Google are extremely good at finding a needle in a haystack. The hierarchies of catalogue portals and Google point to ways in which information can be more efficiently retrieved. Many portals are built from taxonomic hierarchies, in effect metadata-based navigation.

## Data Web

The current Web is primarily a very large number of hyperlinked documents. Whether they're written in loose HTML or more controlled XHTML format, these documents are designed for human reading. The intended path of use goes directly from the organized bits of data through a renderer to the end user. The Web is currently closer to a microfiche repository with an optical viewer than

Figure 2. *The missing Webs (adopted by Ayer, 2004)*



a knowledge representation system. But much of the information on the world's computers isn't in this form; it exists as chunks of information relating to real-world or abstract notions and the relationships between these chunks. As a generalization it could be called relational data, and in fact much of it is stored in quasi-relational SQL databases. But on the current Web, data like this is only usually available through very narrow, human-oriented interfaces. A lot of data held by companies and other organizations will be commercially or politically sensitive, and would need to be kept private. But a considerable proportion of it could be made available more widely to general benefit. Given a framework that supports differing levels of access control, data could be published anywhere in between the private/public extremes.

### **Trusted Web**

If information relating to the source of information can be reliably managed, then this opens up potential in several directions. Being sure of aspects like 'who asserted' and 'when' related to facts enables any conclusions inferred from statements based on those facts to carry some of that assurance.

### **Dynamic Web**

*A visitor from another planet might be forgiven for thinking that computers are solely communication devices. Apart from infrastructure wiring, the Web barely acknowledges that computers are good for computing. To take the computing model beyond the isolated mainframe or desktop PC requires integration of software across organization and even application boundaries.*

### **Transparent Web**

The Web Service approach of passing messages between systems offers a partial solution to making the Web more dynamic. For example, material contained in relational databases can be exposed, so their information becomes as available as that of published documents. But as already noted, the interface tends to be narrow. A database of a hundred tables, a thousand columns, and a million rows may appear on the Web as a single node through which queries have to be tunneled. For efficient interaction between end users and services and between services, a level of transparency is needed in which parcels don't have to be opened to discover their contents.

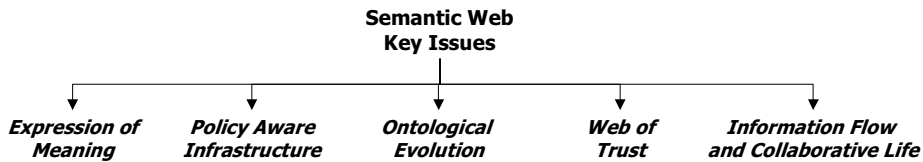
### **Ubiquitous, User-Friendly Web**

Currently most access to the Web takes place through PCs or laptops. There has been some extension into smaller mobile devices, as well as TV-based systems. Wireless has also helped to break some physical restrictions. But still, these are still relatively specialized interfaces; access is far from being on hand everywhere to everyone. In terms of user-friendliness, the Web is generally accessed through a[n] HTML-oriented browser. This usually means read-only access, in very limited single mode of interaction. It lags far behind what is expected of "fat" desktop PC applications. Ubiquity and user-friendliness are key to humanity getting the maximum benefit of the Web, for people to have their abilities augmented at individual and societal levels.

### **Unified Web**

The connectivity of the Web occurs at the level of hyperlinks, in effect the only

Figure 3. Semantic Web key issues



shared languages are fairly low-level protocols. For the Web to be really useful, more sophisticated connectivity is needed. This requires language to describe the entities involved and the relationships between them. Given the scale and diversity of information sources, whatever language is used must be applicable in a very generic way. The only languages that are likely to fit the bill are mathematical, and the prime contenders are understandable in terms of first-order logic.

## SEMANTIC WEB KEY ISSUES

The Semantic Web Activity of the W3C is the key driver for promoting the Semantic Web vision. In our interview with Eric Miller, he outlined four interesting areas:

- **Creating a Policy Aware Infrastructure:** The development of a Policy Aware Infrastructure for the Web is required. The Semantic Web will only achieve its potential as an information space for the free flow of scientific and cultural information if its infrastructure supports a full range of fine-grained policy controls over the information contained in the Semantic Web. If we are going to entrust more of our knowledge to the Semantic Web, we must be assured that the Web will respect many more of the social agreements that we

enforce in the physical world. For the Semantic Web includes not only freely available information, but also personal information and information available to a person or agent only as a result of its membership in groups. A policy-aware infrastructure — one that gives information creators and users the types of control over information we have all become accustomed to in the physical world such as the ability to assert and exercise privacy and intellectual property rights — will make the Semantic Web into a vibrant and humane environment for sharing knowledge and collaborating on [a] wide range of intellectual enterprises.

- **Ontological Evolution:** An important goal of the Semantic Web is to address the problem that, in the course of scientific (or any) endeavor, one changes the vocabularies one uses to organize, discover, and communicate. A given vocabulary may be refined, resulting in a need for migration from old to new. Communication between distinct groups using different vocabularies creates the need to create common vocabularies, which optimally suit all involved. Semantic Web techniques should make this difficult process of creating new common vocabularies as easy as possible. The Semantic Web already removes confusion by giv-



ing each term a globally unique URI. OWL ontologies and rules languages allow relationships between old and new terms to be expressed. There is, however, little experience with the serious management of such evolution. The Semantic Web needs to incorporate versioning and provenance within its foundation. Human understanding changes and statements that we once thought were accurate are later described to be inaccurate. However, the original statement should not be deleted from our corpus of human knowledge. The Semantic Web should not be required to forget that a statement was once believed to be a true statement. Versioning is such a common approach to representing discrete states of understanding that it warrants explicit treatment in the Semantic Web.

- **Web of Trust:** Trust in the human social context is based on constantly evolving and adapting information. Two parties may trust each other based on a history of mutual interaction, based on formal contracts that in turn rely on other established systems (e.g., legal and legislative), and based on risk analysis of a failure of any party to perform as agreed. A trust language for the Semantic Web that is capable of representing these complex and evolving relationships will be crucial to our future ability to build software that behaves more in the manner of an intelligent assistant than a rote rules processor.
- **Information Flow and Collaborative Life:** Many tools used with collaborating groups today instrument the flow of data, information, and knowledge. One of the challenges we will meet is to strike a balance between requiring authors to do more at the outset to make information

machine processable, insisting that everything the machine could use to answer a question be recognized and identified by the (human) questioner, and leaving large quantities of information inaccessible to the machine.

In this list of semantic Web key issues, information systems researchers can find many interesting research topics to contribute. In the interviews, we tried to get feedback on the open research agenda and the “hot topics” that require a multidisciplinary approach. Hendler and Sheth shared with us their thoughts on this issue. More specifically, Hendler’s short list includes five significant areas of research:

1. On the Semantic Web the **ontologies are linked together** and can use terms from other ontologies and change them. The system is open and distributed, and there is no way to guarantee consistency. How do we do reasoning in this kind of distributed and inconsistent system?
2. **Social networks** are becoming very popular on the Web, and it is clear that Semantic Web technologies help support large, distributed networks of people who know other people (like the Friend of a Friend, FOAF, work). What new and exciting things can explicit semantics add to these?
3. How are **traditional** technologies (information retrieval, artificial intelligence, etc.) changed by Web semantics and Semantic Web languages?
4. One of the promises of the Semantic Web is that it will let us **bring databases and structured information sources (like spreadsheets) to the Web**. How will query and search engines for this kind of information work?

## 5. How will semantics function in the emerging world of mobile and ubiquitous computing and other emerging IT trends?

I find Sheth's perspective refreshing because he is among the very few lucky guys who had an opportunity to simultaneously work with the entire span of research, prototyping, technology transfer, commercialization, and real-world application deployment:

*"At LSDIS, I can work with colleagues and a large group of PhD students on long-term and conceptual research which allows me to collaborate with industry and provide inputs to standards activities. We have twice licensed technology resulting from our research, leading to start ups, including Semagix (earlier Taalee). At Semagix, I get to work with smart engineers — some of whom are LSDIS alumni — to develop a leading product in SW and architect customer specific solutions. On the same day I can work on research papers and prototypes, as well as deal with challenges of a deployment at a Fortune 500 customer. It has been incredibly exciting."*

This is not the only reason why I respect his opinion, but this mix of activities is enough to trust Sheth as a leader. So I asked him to sketch for me his hot topics in Next-Generation Web Research:

*"It's hard to pick a few, but here I have a few favorite ones."*

**In [the] research arena,** these include:

1. increasing automatic extraction/annotation of newer forms of digital media, including streaming media, broadcast TV, and sensor-generated data streams;
2. complementing semantic or thematic metadata (and corresponding domain ontologies) with spatial and temporal metadata and ontologies, and providing

comprehensive spatio-temporal thematic reasoning; and

3. extending semantics description of static aspects (such as data input and output) of resources or Web services to descriptions related to functional and execution behavior and quality of service, along with increasing semantic support for dynamic nature of Web processes.

**In [the] commercial and application arena,** some of the favorites include:

4. automated literature search and mining for pharmaceutical R&D;
5. business intelligence applications of opinion and brand management for marketing; and
6. increasing use of semantics in Web search especially as more major players compete with Google.

## THE WAY AHEAD

A lot of questioning is related to the time required for the realization of the Semantic Web. Sheth put his vision for the Semantic Web in an interesting triangle:

*"My view is an amalgamation of what I have seen on 'experiential computing' by Ramesh Jain, 'computing with words' by Lotfi Zadeh, and 'humanist computing' by Jonathan Rossiter. For those focused on semantics and IS, we still need to address the difficult and fundamental problem of identifying entities (from unstructured text), semantic disambiguation, and discovering (potentially fuzzy, inexact, or probabilistic) relationships. And while formal representation and techniques certainly have a role, we need to find [a] much better way for involving humans — much more than in human interfaces and visualization issues — in any approach supporting semantics and knowledge management."*

Hendler contributed to this debate with an to-the-point comment:

*“I think we are going to see continually evolving capabilities based on the Semantic Web infrastructure. I think the first application area where we see it deployed is in enterprise application. That will let us see the creation of ‘islands’ of Semantic Web functionality. We will also see the Semantic Web allowing the creation of easier-to-build and -run Web portals. These will also give us areas of content to link together, and as all these things do link together, the Web of metadata will grow, and we will see the Semantic Web really emerge.”*

Certainly, the Semantic Web cannot be considered as a general milestone or an illusion. A lot of things have to be done, and this road ahead seems to require a well-defined step-by-step approach. Bussler put things in perspective:

*“The Semantic Web is a long-term effort working towards a clear goal, not at all changing every year. Solid progress based on real impact creates a successful area, solid as well as healthy growth, and never a bubble. All involved parties — DERI, research groups, industry, standards organizations, and customers — are interested in making the Semantic Web a reality, not a fashion or a bubble at all. Milestones going forward will be industry-wide pick-up of Semantic Web and Semantic Web Services technology, broad application in all industrial domains, and an ongoing establishment of Semantic Web and Semantic Web Services research groups in universities and research institutes worldwide.”*

So the obvious question is, where are we now? Miller commented on the success of the WWW2004 Web Conference:

*“The WWW2004 Web Conference had a huge Semantic Web focus that permeated almost all aspects of the conference. The energy at the meeting, the collaboration occurring in the*

*corners and throughout the night, reminded me of the second Web conference in Chicago. In Chicago, it seems to me this was a turning point, as everyone who attended realized the Web was not a fad, but rather something that was going to revolutionize how we communicate. The WWW2004 Conference had a similar impact on me with regards to the Semantic Web. The technologies and toolkits are maturing. Semantic Web applications are becoming far more prevalent. Novel ideas for how these technologies may be used are happening on a daily basis. It was quite a week!”*

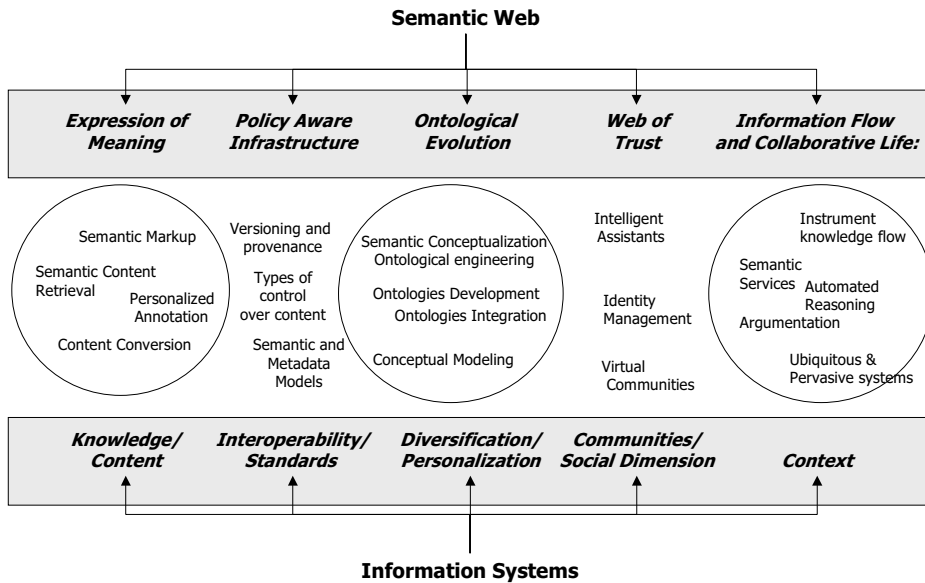
## THE SEMANTIC WEB EQUATION IN IS

This article is not a research paper; rather, it is a synthesis of opinions, ideas, and thoughts. I decided to draw a line and to summarize my understanding of the Semantic Web and its role in information systems from a naïve’s perspective. In Figure 4, key Semantic Web issues are combined with some ontological perceptions for information systems. This rich picture of Semantic Web and information systems research agendas can be a useful guide for putting ourselves within the new context of Semantic Web-enabled information systems. In the three cyclical areas, we can see the basic research streams and topics that currently gain the main interest of researchers. In fact five pairs describe converging actions of semantic Web and information systems research:

- Expression of meaning/managing of knowledge content
- Ontological evolution/diversification-personalization
- Information flow and collaborative life/context
- Policy aware infrastructure/interoperability-standards
- Web of trust/communities—social dimension



Figure 4. Semantic Web/information systems landscape



## ALSO IN THIS ISSUE

In this first issue we tried to have a balanced mix of articles. The second article of the issue, by Rahul Singh, Lakshmi Iyer, and A.F. Salam, co-organizers of the Semantic E-Business Track in AMCIS and guest editors of a special issue on the same topic in *Communications of the ACM*, unfold the research agenda of “Semantic E-Business.” They present a holistic view of semantic e-business that integrates emergent and well-grounded Semantic Web technologies to improve the current state of the art in the transparency of e-business processes.

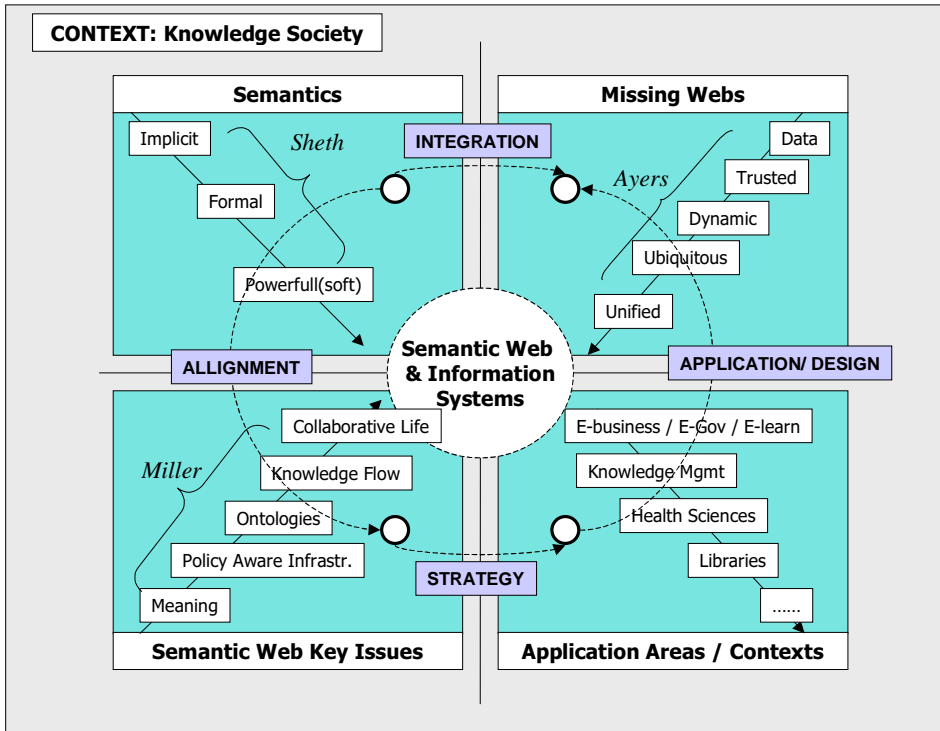
Given the great interest on ontologies development and tools that facilitate the relevant process, we decided to include in the inaugural issue two papers that explain some ontological engineering considerations. Of course we must from the beginning distinguish the nature of our journal. We are not seeking to provide to the research commu-

nity a solid technical journal. Instead, we want to pay more attention to the business issues and the drivers of applied technologies, and from this perspective we emphasize the discussion of research problems, the business justification, and the new facts of the Semantic Web towards real-world problems.

In the third article of this inaugural issue, Aditya Kalyanpur, Bijan Parsia, and James Hendler, from the University of Maryland, describe “A Tool for Working with Web Ontologies.” Beginners in the SW will find this article very interesting since they will see how ontologies and the Semantic Web affect the way we structure and exploit knowledge. Without emphasis on research findings, this article is a good starting point for people interested in working with Web ontologies.

The last article of this issue has a similar orientation. Artem Chebotko, Yu Deng, Shiyong Lu, Farshad Fotouhi, and Anthony Aristar present “An Ontology-Based Multimedia Annotator for the Semantic Web of

Figure 5. Semantic Web definition for information systems



Language Engineering” for multimedia linguistic data. This type of research, I do believe, in the next year will be a must for SW evolution. The management of multimedia content lacks from several perspectives and ontological insights, and SW technologies provide new insights to the problem.

I conclude this editorial with a figure of the knowledge society I would develop in order to communicate my understanding of the Semantic Web. In a way this is for me the translation of the initial equation: “What is the difference in comparison to the WWW that we all know?”

The Semantic Web in the context of information systems research is “*the integration of semantics in the context of ‘missing Webs’ towards the alignment and proposition of new strategies that capital-*

*ize on semantic Web key issues and provide value in specific information systems contexts.*”

This general definition is the mission statement for our journal. In the forthcoming issues we plan very interesting things. We decided to follow the difficult journey of planning and organizing a new journal from scratch in a very short time. In this journey we want your help; we are looking forward for your comments, we need your participation in AIS SIGSEMIS and SIGODIS activities, and moreover we are open for ideas on collaboration. See you in the next issue.

## ACKNOWLEDGMENTS

I would like personally to thank the Council of AIS for their great support and

especially Ritu Agarwal. Without their approval our efforts could not find such a creative place to foster growth.

Moreover I would like to thank Gottfried Vossen and Gerd Wagner for their warm participation in the initial proposal of SIGSEMIS.

My deepest appreciation to all the renewed academics and practitioners who joined our SIGSEMIS board and also provide the Editorial Board of *IJWSIS*.

Special thanks and great respect to Professor Amit Sheth, who is kindly providing inspiration for our current strategy and the future of our activities. Without his help, this journal could not be a reality.

Many thanks also to AIS SIGSEMIS members for their collaboration towards our objective for cultivating the Semantic Web vision in the information systems research community. I invite you to join our SIG and be part of an exciting community. ([www.sigsemis.org/sig/membership/document\\_view](http://www.sigsemis.org/sig/membership/document_view)).

I am grateful to Mehdi Khosrow-Pour, Jan Travers, and all the personnel at Idea Group Inc. who do their best for serving our journal.

My last GREAT THANKS to my wife Theodora, and my little children Dimitris and Hara, who see me every night after midnight authoring and editing important papers.

## REFERENCES

- Ayers, D. (2004). The missing Webs. *AIS SIGSEMIS Bulletin*, 1(2), 18-23. Available at [www.sigsemis.org](http://www.sigsemis.org)
- Bussler, C. (2004). An interview with Christoph Bussler. *AIS SIGSEMIS Bulletin*, 1(2), 74-80. Available at [www.sigsemis.org](http://www.sigsemis.org)
- Miller, E. (2004). An interview with Eric Miller. *AIS SIGSEMIS Bulletin*, 1(2), 3-10. Available at [www.sigsemis.org](http://www.sigsemis.org)
- Hendler, J. (2004). An interview with James Hendler. *AIS SIGSEMIS Bulletin*, 1(1), 2-4. Available at [www.sigsemis.org](http://www.sigsemis.org)
- Sheth, A. (2004). An interview with Amit Sheth. *AIS SIGSEMIS Bulletin*, 1(1), 14-19. Available at [www.sigsemis.org](http://www.sigsemis.org)